



Lisbon School  
of Economics  
& Management  
Universidade de Lisboa



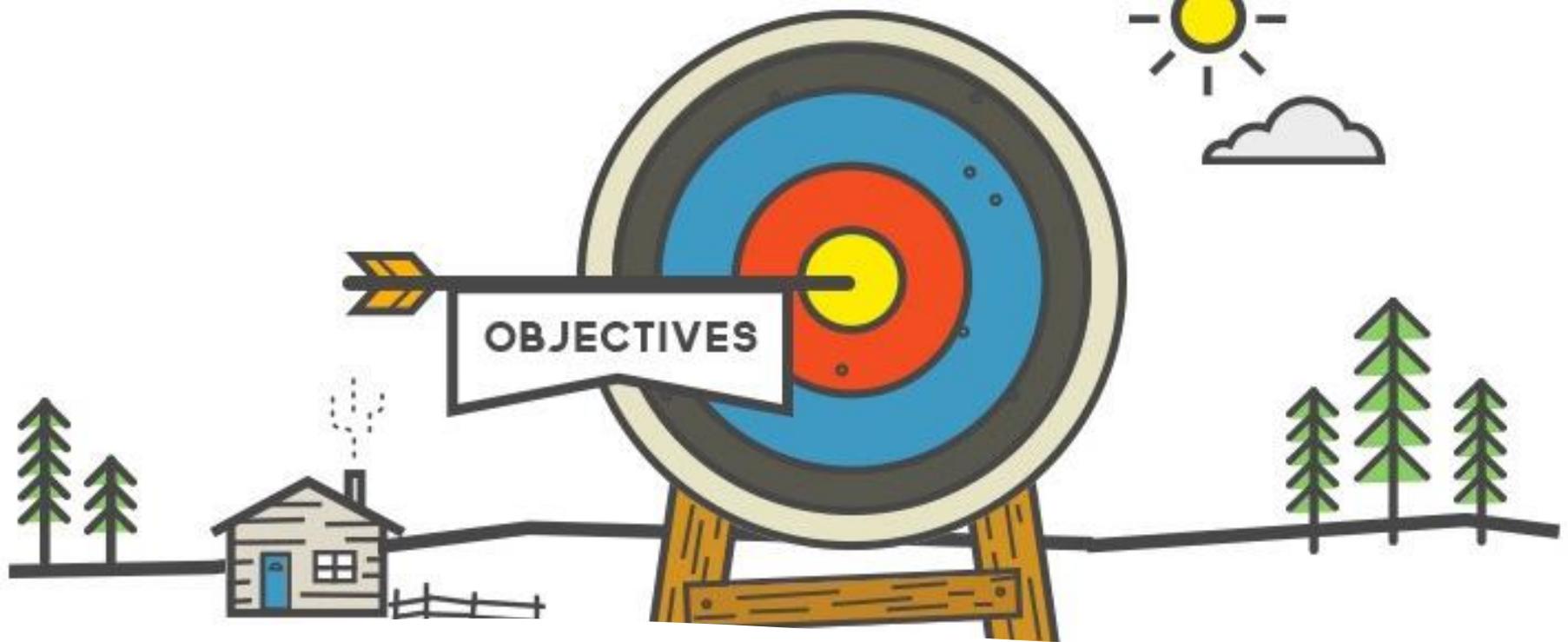
# Introduction to Data Visualization

Prof. Carlos J. Costa, PhD

Saeed Angorani, DBA

# Session Overview

- Importance of Data Visualization in Business Analytics
- History and Evolution of Data Visualization
- Cognitive Aspects of Visual Perception: How we process and interpret visual data
- Key Principles of Effective Data Visualization (Clarity, Accuracy, Efficiency)
- Common Mistakes and Misleading Visuals



## Learning Goals

By the end of this session, you will be able to:

1. Articulate the strategic value of data visualization in business decision-making and communication
2. Trace the historical evolution of data visualization from early statistical graphics to modern interactive dashboards
3. Explain the cognitive mechanisms underlying visual perception and how they influence data interpretation
4. Apply the core principles of clarity, accuracy, and efficiency when designing or evaluating visualizations
5. Identify and avoid common pitfalls that lead to misleading or ineffective visuals

# Why Visualization Matters

- **The Human Brain Processes Images 60,000× Faster Than Text**
- 90% of information transmitted to the brain is visual
- We can identify visual patterns in milliseconds
- A well-designed chart can replace pages of tables

*"The greatest value of a picture is when it forces us to notice what we never expected to see."* — John Tukey

# Data Overload is Real

## The Modern Business Challenge

Year	Global Data Created
2010	2 zettabytes
2020	64 zettabytes
2025	180+ zettabytes (projected)

Without visualization, this data remains noise, not signal.

# Visualization as a Business Tool

## Four Core Functions in Organizations

- Exploration - Discover patterns and anomalies
- Explanation - Communicate findings to stakeholders
- Monitoring - Track KPIs in real-time dashboards
- Persuasion - Support decisions and drive action

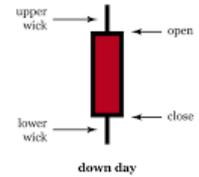
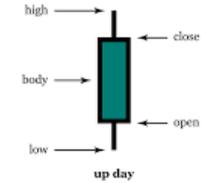
# Business Impact Examples

## Real-World Applications

- **Retail:** Heat maps of store traffic optimize product placement
- **Finance:** Candlestick charts reveal market sentiment
- **Healthcare:** Patient flow dashboards reduce wait times by 20-30%
- **Marketing:** Funnel visualizations identify conversion bottlenecks
- **Supply Chain:** Geographic maps track logistics in real-time



Candlestick chart



Analyzing patient wait time for emergency rooms



# The Cost of Poor Visualization

- **When Visuals Fail, Decisions Fail**
- Misread charts lead to incorrect forecasts
- Cluttered dashboards cause "analysis paralysis"
- Misleading graphs erode stakeholder trust

# History and Evolution of Data Visualization

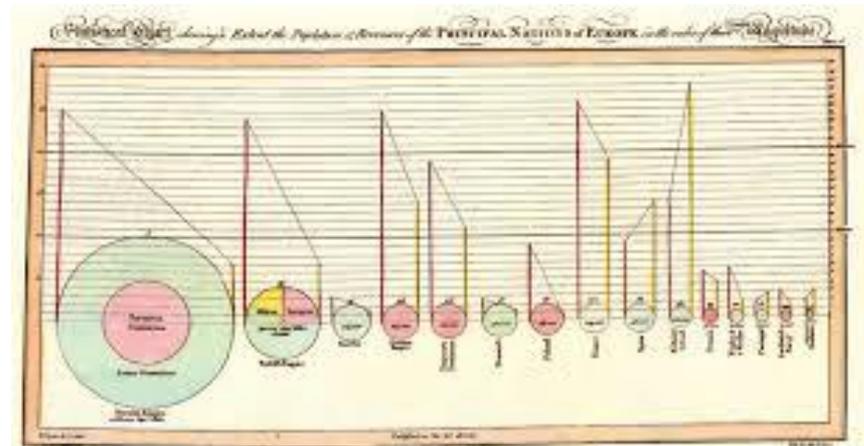
- **Early Beginnings (Pre-1800)**
- **The Golden Age (1750-1900)**
- **Florence Nightingale (1858)**
- **Charles Joseph Minard (1869)**
- **The Dark Ages (1900-1970)**
- **The Renaissance (1970-Present)**

# Early Beginnings (Pre-1800)

- **The Roots of Visual Data**
- **~6000 BCE:** Earliest known maps (Çatalhöyük, Turkey)
- **~3000 BCE:** Egyptian land surveys and astronomical charts
- **1350:** Nicole Oresme's proto-bar charts for velocity
- **1644:** Michael van Langren's first known statistical graph (longitude estimates)

# The Golden Age (1750-1900)

- **Pioneers Who Shaped the Field**
- **William Playfair (1759-1823)** — "Father of Statistical Graphics"
- Invented the line graph (1786)
- Invented the bar chart (1786)
- Invented the pie chart (1801)





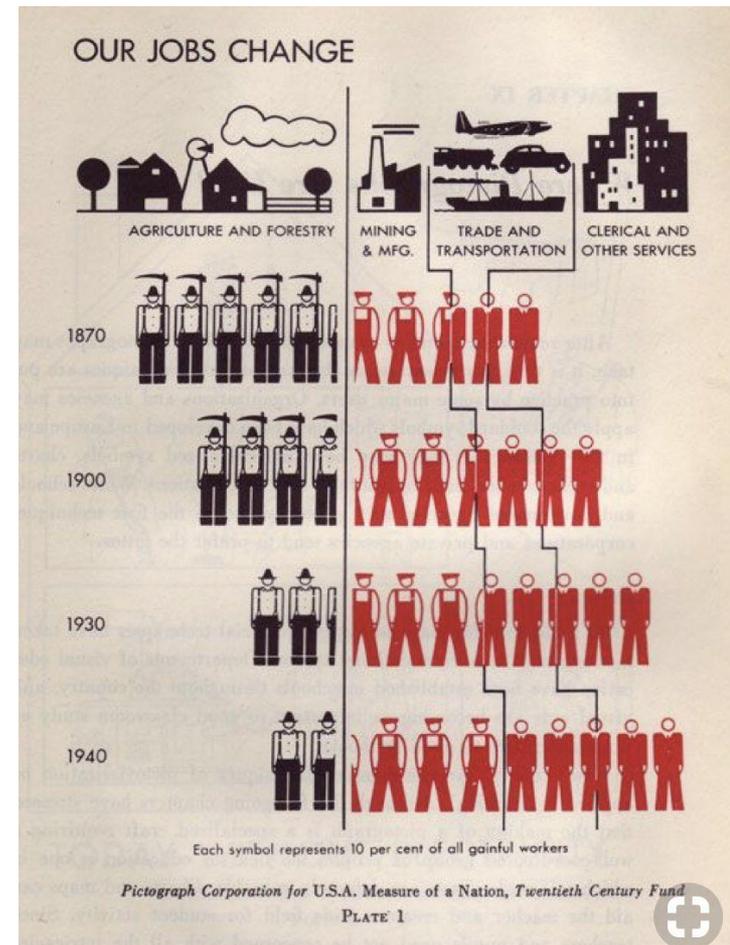


# The Dark Ages (1900-1970)

## A Period of Stagnation

- Statistics became more mathematical, less visual
- Academic focus shifted to formulas and tables
- Visualization seen as "unscientific" or merely decorative

Notable exception: **Otto Neurath's ISOTYPE** (1920s-30s)  
— pictorial statistics for public education



# The Renaissance (1970-Present)

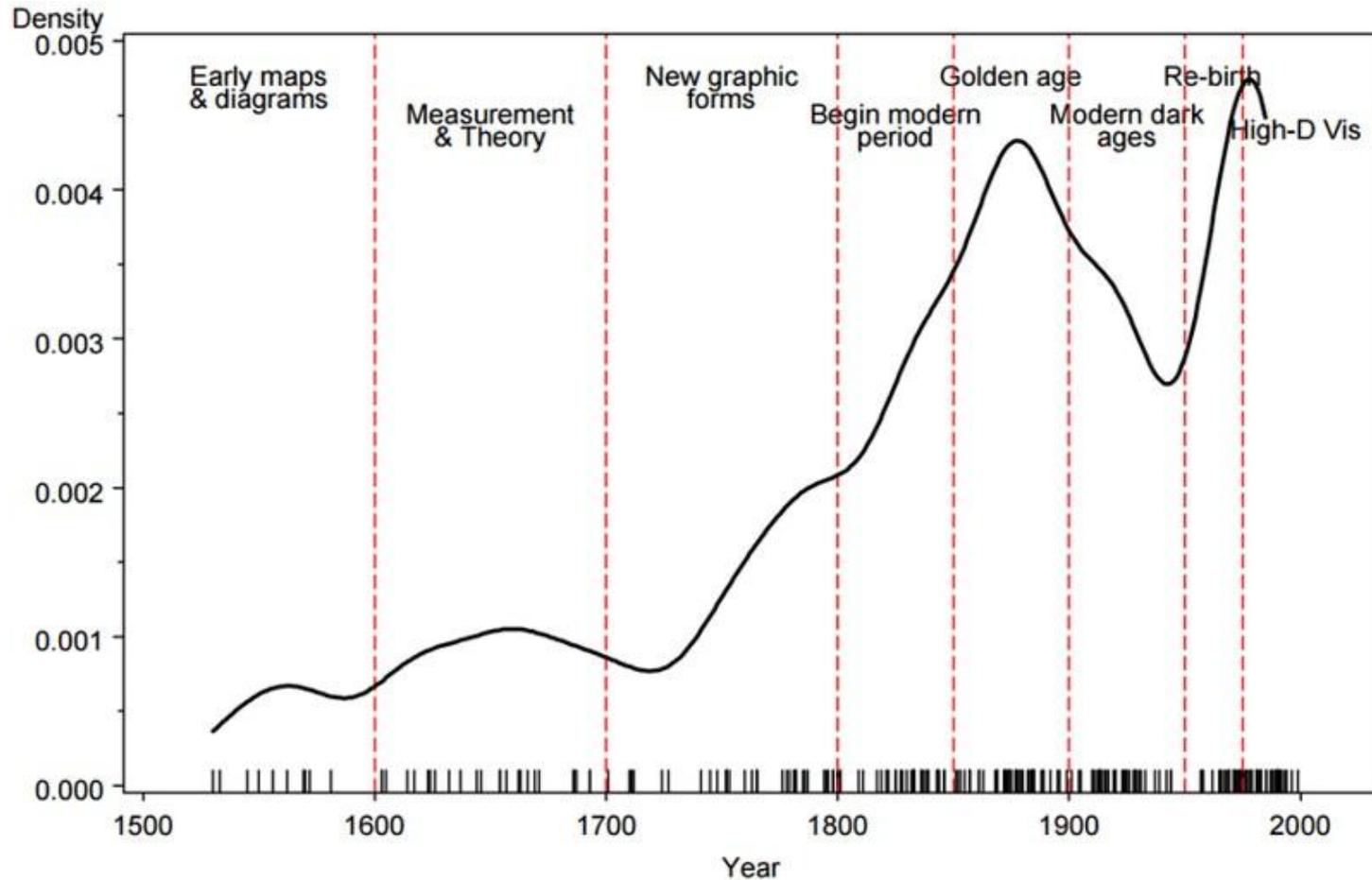
- **Technology Enables a New Era**

Decade	Key Developments
1970s	John Tukey's Exploratory Data Analysis; first computer graphics
1980s	Edward Tufte's foundational books; spreadsheet software
1990s	Interactive visualization; early web graphics
2000s	Tableau, D3.js; "Big Data" visualization challenges
2010s	Mobile dashboards; real-time streaming data
2020s	AI-assisted visualization; immersive/VR data experiences

# Key Figures to Know

- **The Modern Canon**
- **John Tukey** — Exploratory Data Analysis, box plot inventor
- **Edward Tufte** — "The Visual Display of Quantitative Information"
- **Jacques Bertin** — "Semiology of Graphics" (visual variables)
- **Ben Shneiderman** — Information visualization, treemaps
- **Hans Rosling** — Animated bubble charts, Gapminder
- **Alberto Cairo** — "The Truthful Art," journalism focus

# Timeline Summary



# Cognitive Aspects of Visual Perception

- **How We See Data**
- The Science of Visual Perception



# Two Systems of Thinking

- Daniel Kahneman's Framework

System 1	System 2
Fast, automatic	Slow, deliberate
Effortless	Requires concentration
Pattern recognition	Logical analysis
Emotional	Rational

# Pre-Attentive Processing

- **What We See Before We "Think"**
- Certain visual attributes are processed in under 250 milliseconds:
  - **Color** (hue, intensity)
  - **Form** (size, shape, orientation)
  - **Position** (spatial location)
  - **Motion** (flicker, movement)

# Pre-Attentive Attributes

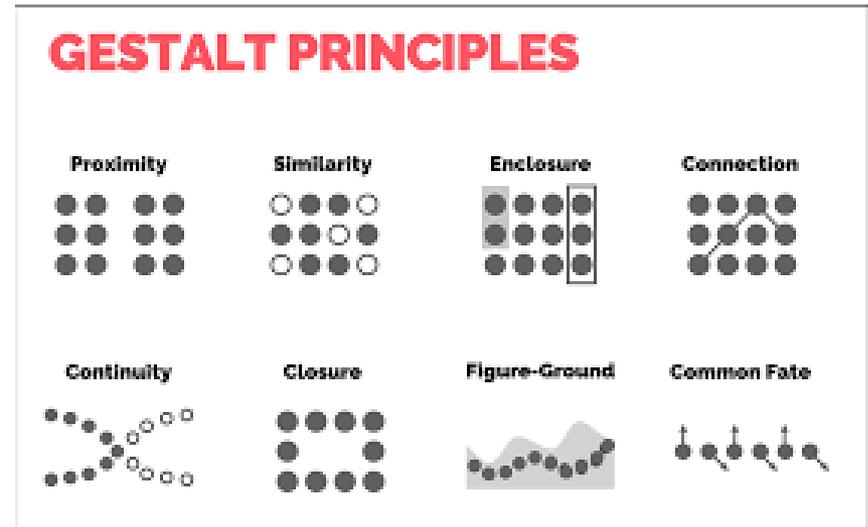
Attribute	Best For
Color hue	Categories (max 6-8)
Color intensity	Magnitude, emphasis
Size	Quantitative comparison
Position	Precise comparison
Orientation	Direction, categories
Shape	Categories (limited)

# Gestalt Principles

## How We Group Visual Elements

Our brains automatically organize visual information:

- **Proximity** — Objects close together seem related
- **Similarity** — Similar objects seem grouped
- **Enclosure** — Bounded areas form groups
- **Continuity** — We follow smooth lines/curves
- **Connection** — Connected elements seem related
- **Closure** — We complete incomplete shapes



# Gestalt in Practice

- **Proximity:** Grouped bar charts
- **Similarity:** Color-coded data series
- **Enclosure:** Background shading for regions
- **Continuity:** Line charts
- **Connection:** Network diagrams
- **Closure:** Implied axes and gridlines

# Working Memory Limits

## The Magical Number $7 \pm 2$

- George Miller (1956): We can hold approximately 7 items in working memory.

## Implications for visualization:

- Limit legend categories to 5-7
- Avoid charts with 15 different colors
- Use hierarchy to chunk information
- Don't force users to remember across views

# Change Blindness

## We Miss More Than We Think

- Users may not notice changes between dashboard refreshes
- Important alerts need strong visual signals
- Animation can help draw attention to changes

# Perceptual Accuracy by Chart Type

## Cleveland & McGill's Hierarchy (1984)

Most accurate to least accurate position judgments:

1. Position along a common scale (bar chart)
2. Position along non-aligned scales
3. Length
4. Angle/Slope
5. Area
6. Volume
7. Color saturation/density

**Practical implication:** Bar charts beat pie charts for precise comparison.

# The Pie Chart Problem

## Why pie charts often fail:

- Humans are poor at comparing angles
- Difficult with more than 3-4 slices
- Impossible to compare across multiple pies
- 3D pie charts make everything worse

**When pie charts work:** Showing parts of a whole with 2-3 dominant segments, when precision isn't critical.

# Color Perception Challenges

## Designing for All Users

- ~8% of men and ~0.5% of women have color vision deficiency
- Red-green colorblindness is most common
- Never rely on color alone to convey meaning

## Solutions:

- Use redundant encoding (color + shape, color + label)
- Test with colorblindness simulators
- Use colorblind-friendly palettes (e.g., viridis, ColorBrewer)

# Cognitive Load Theory

- **Every Element Costs Mental Effort**

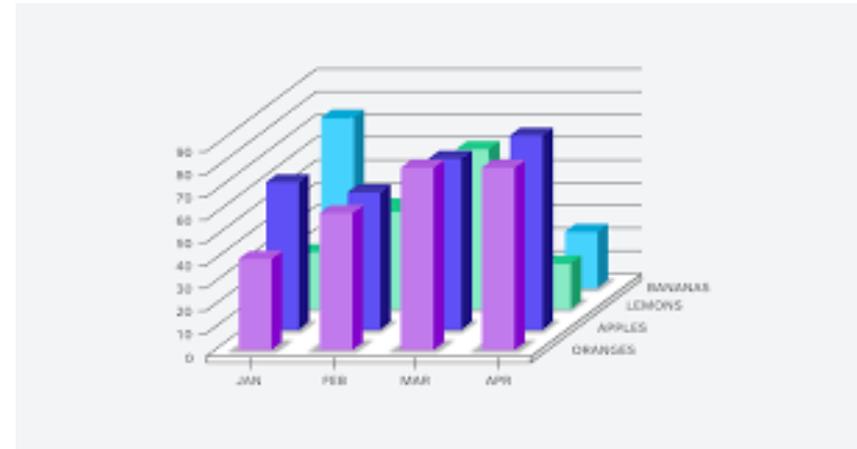
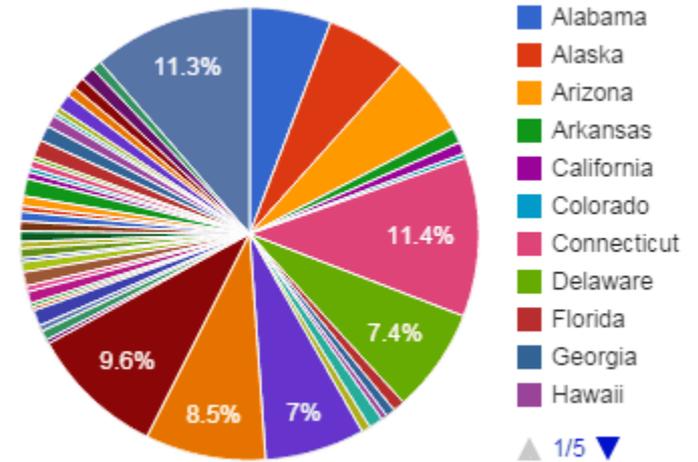
Load Type	Description	Design Response
Intrinsic	Complexity of the data itself	Simplify where possible
Extraneous	Poor design adding confusion	Eliminate chart junk
Germane	Effort spent learning/understanding	Support with clear labels

- **Goal:** Minimize extraneous load so users can focus on the data.

# Exercise: Spot the Cognitive Issues

**In pairs, identify:**

- Pre-attentive attributes being misused
- Gestalt principles being violated
- Unnecessary cognitive load



# The Three Core Principles

**Every good visualization balances:**

- 1. Clarity** — The message is immediately understandable
- 2. Accuracy** — The data is represented truthfully
- 3. Efficiency** — Maximum insight with minimum ink

These principles sometimes tension with each other - design is about finding the right balance.

# Principle 1: Clarity

## "Don't Make Me Think"

A clear visualization:

- Has an obvious main message
- Uses familiar chart types appropriately
- Labels everything the user needs
- Removes ambiguity
- Works without explanation

**Test:** Can someone understand the main point in 5 seconds?

# Clarity: Choose the Right Chart

Data Relationship	Recommended Charts
Comparison	Bar chart, dot plot
Trend over time	Line chart, area chart
Part-to-whole	Stacked bar, treemap, (pie with caution)
Distribution	Histogram, box plot, violin plot
Correlation	Scatter plot, bubble chart
Geographic	Choropleth map, symbol map

# Clarity: Labeling Best Practices

## Labels Should Reduce, Not Add, Cognitive Load

- ✓ Direct labeling on data points (when space permits)
- ✓ Clear, descriptive titles that state the insight
- ✓ Axis labels with units
- ✓ Annotations for key events or outliers
- ✗ Legends that require back-and-forth eye movement
- ✗ Cryptic abbreviations
- ✗ Missing units or context

# Principle 2: Accuracy

## "Above All, Do No Harm"

An accurate visualization:

- Represents data proportionally
- Starts axes at appropriate values
- Doesn't cherry-pick time ranges
- Shows uncertainty when relevant
- Provides context for interpretation

**Trust is hard to build and easy to destroy.**

# Accuracy: The Baseline Problem

## Same data, different impressions:

- Truncated axes exaggerate differences
- Sometimes justified (e.g., stock prices), but must be clearly labeled
- Default: Start bar charts at zero

# Accuracy: Aspect Ratio Matters

## William Cleveland's "Banking to 45°"

- Line slopes should average around 45° for optimal perception
- Too wide = trends look flat
- Too tall = trends look dramatic

Choose aspect ratios that represent change fairly

# Accuracy: Show Uncertainty

## Data is Rarely Certain

Ways to represent uncertainty:

- Error bars
- Confidence intervals
- Gradient/blur effects
- Ensemble plots (multiple scenarios)
- Explicit annotations

**Hiding uncertainty is a form of dishonesty.**

# Principle 3: Efficiency

## Edward Tufte's Data-Ink Ratio

$$\text{Data-Ink Ratio} = \frac{\text{Ink used for data}}{\text{Total ink used}}$$

### Maximize this ratio by removing:

- Unnecessary gridlines
- Redundant labels
- Decorative elements ("chart junk")
- 3D effects that add no information
- Excessive colors

# Key Misleading Visuals & Mistakes

- **Truncated Y-Axis (Non-Zero Baseline):** Starting the Y-axis at a value other than zero exaggerates minor differences, making small changes appear massive.
- **Improper Chart Type:** Using pie charts for time-series data or bubble charts for precise comparisons misrepresents the intended story.
- **Misleading Colors & Contrast:** High-contrast colors can falsely imply significant differences between data points, while too many colors create noise.
- **3D Graphics & Distortion:** Using 3D charts makes it hard to compare segments accurately, leading to cognitive distortion of data.
- **Cluttered Design & Overloading:** Too many data points, labels, and irrelevant elements make it difficult for viewers to identify the main message.
- **Missing Context/Annotations:** Lacking context (e.g., no titles, axis labels, or units) results in interpretation errors.

# How to Avoid Misleading Visuals

- Always start the Y-axis at zero when dealing with bar charts to ensure accurate comparison.
- Simplify the design, removing unnecessary 3D effects, shadows, or excessive colors.
- Select the appropriate chart for the data type (e.g., line charts for trends, bar charts for comparison).
- Use consistent, meaningful colors that enhance understanding rather than causing confusion.
- Provide clear titles and labels that explain exactly what the data represents

# Key Takeaways

## Five Things to Remember

1. **Visualization is strategic** — it transforms data into decisions, not just pictures
2. **Design for the brain** — leverage pre-attentive processing; respect cognitive limits
3. **Three principles guide everything** — Clarity, Accuracy, Efficiency
4. **Truthfulness is non-negotiable** — truncated axes, cherry-picked data, and 3D effects destroy trust
5. **Less is more** — if it doesn't serve the data, remove it

# Before You Share Any Visualization

## The 5-Second Test

- ✓ Main message clear in 5 seconds?
- ✓ Data represented truthfully?
- ✓ No unnecessary elements?
- ✓ Works for colorblind users?
- ✓ Would you trust it if someone else made it?

**If any answer is "no" - revise.**

# Closing Thought

**"The purpose of visualization is insight, not pictures." Ben Shneiderman**

Your job: be a translator between data and human understanding.

**Questions?**

# Questions

- Thank you!
- Questions and discussion.